DOCUMENT RESUME

ED 261 079 TM 850 481

AUTHOR Harnisch, Delwyn L.
TITLE Computer Application Issues in Certification and Licensure Testing.
PUB DATE Mar 85
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Information Analyses (070)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Adaptive Testing; *Certification; *Computer Assisted Testing; Cost Effectiveness; Individual Testing; Item Analysis; Item Banks; Latent Trait Theory; Mastery Tests; *Occupational Tests; Test Construction; Testing Problems; Test Items; Test Length; Test Reliability; *Test Validity

ABSTRACT

Computer adaptive testing systems are feasible for certification and licensure testing. This is in part due to the availability of extensive yet inexpensive computers. Modern item response theory, combined with computerized adaptive testing, yields a powerful new method of testing which provides greater accuracy and efficiency and less boredom for the examinee. The computer presents each item, scores it, and then selects the next item which is appropriate for the individual examinee; thus, individual testing is much more feasible than it would be using human examiners. There are additional advantages: acceptable accuracy can be achieved with fewer items; more accurate estimates can be obtained at the extreme ends of the ability continuum; test security is improved; tests can be administered on demand when needed; and a greater variety of item types may be included. With licensing and certification tests, it is important to obtain a good pool of test items, to be accurate near the minimum cutting score, to have appropriate unidimensionality, and to use computers in a comfortable environment. It is also useful to have a large number of examinees, approximately 1000 or more. Cost estimates are encouraging, especially when the computer systems are used often. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED261079

Computer Application Issues in Certification and Licensure Testing

Delwyn L. Harnisch

University of Illinois at Urbana-Champaign

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. L. Harnisch

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the annual meeting of the American Educational Research Association, Chicago, March 31 - April 4, 1985.

Note: This paper has been submitted to ERIC and should be available from the ERIC system by June, 1985.

IM 850 481

Computer Application Issues in Certification

and Licensure Testing

Introduction

In the United States approximately 800 occupations are regulated by state governments. In addition, credentials are issued by nongovernmental agencies in hundreds of other areas, making the regulation activity quite widespread. What is clear from all this activity is that regulation either in the form of licensure ("Practice Acts") or certification ("title control") is prevalent throughout the United States (Shimberg, 1981).

Since the original Sunset law was enacted in Colorado in 1976, about two-thirds of the states have established some form of Sunset legislation. As a result, Sunset hearings have led to reforms in some licensing procedures, greater public attention given to the process, and greater awareness of the realities of litigation (Linn, in press). In the last few years, the licensing and certification boards have come to understand the importance of sound test construction and the need for well-documented validation studies. The publication of the Standards for Educational and Psychological Tests is a very useful document because of the guidance it provides to those designing licensure and certification tests. The professional test standards aid in the protection of the rights of the public and the examinee.

The purpose of this paper is to describe some of the computer application issues in certification and licensure testing. The availability of relatively inexpensive computer technology has created many new possibilities for testing in certification and licensure. Computer adaptive tests and simulation tests are just two examples which can readily be implemented with the present technology. Given the time limits for this talk today, I will focus my comments on the computer applications issues for the computer adaptive testing systems.

Large-scale ability testing has been a cornerstone of American education for most of the twentieth century. Information from this activity has aided decision makers in measuring the impact of schools, evaluating the effectiveness of educational intervention programs and achieving other purposes which form the basis of the American meritocracy.

During the past decade substantial criticism has been made of the testing industry. One can say that some of it was deserved but the majority of it was due to poor understanding of the technical issues involved. Since the criticisms were made public, the testing industry has made considerable effort to explain the details of the new advances in testing.

Technological Revolution in Testing

The availability of extensive inexpensive computing opens many new possibilities for testing. Some have referred to this as a technological revolution in testing (Wainer, 1983). In a paper and pencil environment, it is often found to be too cumbersome to administer adaptive tests and simulation tests. With the recent technological advances, the Computerized Adaptive Test (CAT) has become quite feasible. It is worth noting at this time that it takes much more than having improved tools to make a better product. Nonetheless, the possibilities are appealing.

Modern Item Response Theory (IRT) combined with CAT yields a powerful new measuring instrument that permits testing with greater accuracy and efficiency and with less boredom and/or frustration for the examinees. Possibilities for cheating are reduced, appropriateness of the test for various groups of examinees can be investigated and new doors for ability testing in certification and licensure testing can be opened.

Advances in IRT

The recent advances in IRT have provided the possibilities for efficient testing activities. Some of these advances in adaptive testing are recognized

and outlined below.

The logic of adaptive testing, tailoring a test to suit the individual person tested, goes back to the initial Binet intelligence tests which were administered individually. Adaptive tests are also known as tailored tests, branched tests, sequential tests and response-contingent tests.

Given the present day activity of large group testing, such individual attention is impractical. Interactive computers can now replace the human test administrator. The computer can now make decisions many times faster and more accurately than a human; thus making it possible for individualized testing to be attractive again.

These advances in testing technology make no guarantees for improvements in testing. For example, the use of a high-speed optical scanner does not make a test more valid than if the test were scored by hand with the aid of old-fashioned template key. Computerized test administration by itself does not enhance validity. To realize improvements in validity will depend on the quality of the test content and on the nature of the test items that are created for computer administration.

Adaptive tests are presently a part of a program used by the U. S. Armed Services for their primary selection and classification testing. Green, Bock, Humphreys, Linn and Reckase (1984) summarized the basic procedures for adaptive testing as follows:

> Adaptive testing requires a digital computer to present each test item, score each response, and then select the next item that will be most appropriate for the candidate. Adaptive testing seeks to present only items that are appropriate for the test taker's estimated level of skill or ability. Questions that are too easy or too difficult for the candidate contribute very little information about that person's ability. More specifically, each person's first item on an adaptive test generally has about medium difficulty for total population. Those who answer correctly get a harder item; those who answer incorrectly get an easier item. After each response, the examinee's ability is estimated, along with an indication of the accuracy of the estimate. The next item to be posed is one that will be especially informative for a person of the estimated ability, which generally

> means that harder questions are posed after correct answers and easier questions after incorrect answers. The change in item difficulty from step to step is usually large early in the sequence, but becomes smaller as more is learned about the candidate's ability. The process continues until there is enough information to place the person on the ability scale with a specified level of accuracy, or until some more pragmatic criterion is achieved (pp 347-348).

Advantages of CAT

There are numerous advantages of the adaptive testing approach over the paper-and-pencil test. Because of time constraints, I will just discuss a few of these advantages today. For a more detailed list of advantages, see Green's (1983) paper. Some of the advantages of CAT are as follows:

1. A given level of accuracy can be achieved with fewer items.

2. More accurate estimates can be obtained at the extreme ends of the ability continuum.

3. Test security is improved so that cheating or prior test information does not interfere with performance, since each examinee is administered a unique exam.

4. Tests can be administered on demand rather than on a few administration dates each year.

5. Greater variety of item types can be included in the item bank as it need not always be multiple-choice format.

Issues for Certification and Licensure Tests

The methods of adaptive testing and IRT rely upon some strong theoretical assumptions and require substantial information about the characteristics of a large item pool. (For a review of an evaluation plan for CATs, see Green, Bock, Humphreys, Linn & Reckase, 1982). Obtaining a good item pool is often the most difficult task in the development of an adaptive test. If a relatively large pool of items is available from prior administrations of conventional tests along with their response data, then the items can be organized and calibrated for use in the testing system.

Improved measurement at the extremes of ability on a licensure test, although important, is of less importance than the accuracy of measurement near the minimum passing score. The assumption which may cause the greatest problem for some of the licensure and certification exams is that of unidimensionality. Unidimensionality is more critical for adaptive tests than for ability tests (Green, Bock, Humphreys, Linn, & Reckase, 1984). For a review of various ways to obtain evidence of unidimensionality see Hambleton & Murray (1983). Although the assumption is necessary for present applications of IRT, it is not reasonable for a licensure or certification test that was designed to measure several dimensions of professional competence. Where these multiple item domains exist, one solution would be to classify the item pool into relatively homogeneous subsets and then construct adaptive tests for each. Decision rules for certification or licensure could then be based on the scores resulting from each of the adaptive tests.

Some of the certification and licensure programs are run for relatively small populations. For these programs, the ability to have calibrated items may be a serious hindrance to adaptive testing since sample sizes of approximately a thousand or more examinees are needed to adequately estimate the item parameters for use in adaptive testing. On the other hand, programs with large enrollments in their licensure and certification activities would be able to support an adaptive testing system.

There is an apparent mismatch between the improved accuracy provided by the adaptive testing strategy and the goal of the most valid and reliable pass-fail decisions for a licensing test which suggests a need for other adaptive strategies (Linn, in press). In most licensing test situations the concern is with the maximum information to assist in the making of the pass-fail decision. In these situations what is needed is an adaptive strategy that minimizes the errors of classification around the cut point.

It has been demonstrated by Linn, Rock and Cleary (1972) and Owen (1975) that with CAT the average testing time, using the simple sequential testing procedure, can be reduced. A decision is made as to whether another item should be given to the examinee after the administration of each item. Results from these studies indicate that good precision in measuring ability is attained with about 25 items in most cases. Compared to the conventional tests, CAT administrations typically require approximately one-half as many items on the average to reach the same level of accuracy. Additional savings in number of items administered could vary as one uses more advanced procedures proposed by Lord (1980).

Practical Considerations

Some fears have been expressed by many adults, who might be intimidated by taking a test on the computer. Other factors to consider include:

1. Quiet environment free of distractions with a terminal cubicle desirable.

2. Clarity of text presentation on the monitor is essential.

3. There should be centrally located response buttons.

4. There should be the same response from key entry to the next item presentation.

5. Computer systems should not add variability in testing time and presentation of test items.

6. Rest periods should be provided once or twice during a testing session.

7. Presentation of directions and checking for understanding of instructions deserves careful attention in a computer setting.

8. Computer testing systems should allow for monitoring individual progress on the terminals.

Cost Effectiveness

Finally, some of you are probably asking yourselves whether these systems

are cost effective. This is a complex question which is difficult to answer completely. Typically one considers the following three things to evaluate cost effectiveness: cost of test development, cost of measurement precision, and the cost of the computer system (Vale, 1984). In general, the time required to develop materials for adaptive tests is much shorter than that for conventional paper-and-pencil tests where, for security reasons, a new test must be developed for each administration date. Test security is improved with adaptive tests as most of the test redevelopment costs are saved. Adaptive tests can also provide greater measurement precision and domain coverage in typically half the time. No longer will answer sheets be lost or not scanned correctly. The test can be scored immediately, providing possibilities of giving immediate diagnostic information back to the examinee. In general, the costs of the computer system for use only two or three days a year would be rather expensive. However, if the testing program is provided on a year-round basis, the utilization would increase and the costs for the computer system would decrease.

A Look Into the Future

Computer applications in licensure and certification testing will be increasingly used in future years. Lloyd (1984) made the following comments regarding the licensure and certification in the future.

> Physician licensure and certification, or their twenty-first century equivalents, will not be awarded for life; rather they will be achieved when performance reaches the standard and, maintained or not, based on the physician's accumulating record of performance. Standards will be national for certification purposes, state-wide for licensure purposes, local for reimbursement purposes, and personal for self-assessment purposes. Knowledge and awareness of these standards at any time will be available to the physician, as well as the necessary information and expertise to attain it. When a physician's performance continually falls below a standard, diagnostic and remedial programs will be available to guide the physician's performance back to the standard. Consistent performance above the standards will result in honors and awards...(p. 193)

Development of computerized testing systems for licensure and certification involves the integration of measurement and instructional design principles.

Validity of either the licensure or certification exam involves having an exam that assesses the knowledge and skills that are essential for safe and competent performance (Shimberg, 1984). Test development procedures that identify the critical aspects of the occupation, determine the weight given to each aspect of the occupation, and select the items that are fair and designed for the appropriate cognitive level are important elements in this process for implementation on the computer.

Conclusion

The technological revolution in testing has returned us to tailored individualized testing opportunities. The technological revolution along with the advances in modern test theory have made the topics discussed in this symposium a practical option to consider for future licensure and certification testing activities. As a result of available inexpensive computing and the development of a powerful theoretical structure of testing, we can now experience the improved practice of testing.

In summary, I feel that, while many related issues must be taken into consideration by those who wish to develop computer applications for certification and licensure testing, the potential is great for the improvement of testing programs sponsored by licensing agencies and certification groups.

References

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education (Joint Committee, November 1984). Standards for educational and and psychological testing. Washington, D. C.: American Psychological Association.

Green, B. F. (1983) The promise of tailored tests. In H. Wainer & S. A. Messick (Eds.), Principles of modern psychological measurement: A festschrift in honor of Frederic Lord. Hillsdale, NJ: Lawrence Erlbaum Associates.

Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1982). Evaluation plan for the computerized Adaptive Vocational Aptitude Battery. Baltimore, MD: Johns Hopkins University, Department of Psychology.

Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. Journal of Educational Measurement, 21, 347-360.

Hambleton, R. K., & Murray, L. (1983). Some goodness of fit investigations for item response models. In R. K. Hambleton (Ed.), ERIBC Monograph on Application of Item Response Theory. Vancouver: Education Research Institute of British Columbia, 71-94.

Linn, R. L. (in press). Development and trends in licensing and licensure examinations. In J. C. Fortune (Ed.), Understanding testing in occupational licensing: Establishing links between principles of measurement and practices. San Francisco: Jossey Bass, Inc.

Linn, R. L., Rock, D. A., & Cleary, T. A. (1972). Sequential testing for dichotomous decisions. Educational and Psychological Measurement, 32, 85-95.

Lloyd, J. S. (1984). Computers in the future. In J. S. Lloyd (Ed.), Computer Applications in the Evaluation of Physician Competence. Chicago: American Board of Medical Specialties, 191-193.

Lord, F. M. (1980). Application of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associates.

Owen. R. J. (1975). A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. Journal of the American Statistical Association. 70, 351-356.

Shimberg, B. (1981). Testing for licensure and certification. American Psychologist, 36, 1138-1146.

Shimberg, B. (1984). Why validity is THE issue in licensure testing. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Vale, C. D. (1984). The adaptive test as a procedure for assessing competence. In J. S. Lloyd (Ed.), Computer Applications in the Evaluation of Physician Competence. Chicago: American Board of Medical Specialties, 129-134.

Wainer, H. (1983). On item response theory and computerized adaptive tests. Journal of College Admissions. 28, 9-16.

Acknowledgments

The author would like to acknowledge Carolyn Palmer and Michael Levine for their comments on this paper.